Introduction to the D.C. Cancer Registry

Purpose of the Registry

Population-based cancer registries are essential for assessing the extent of cancer burden in a specified geographic area. The District of Columbia Cancer Registry (DCCR) is a population-based cancer registry that collects incidence data on all cancer patients who reside in or who are diagnosed and/or treated for cancer in the District of Columbia. The goals of the DCCR are:

- to determine the incidence of cancer in the District of Columbia with respect to geographic, and demographic characteristics
- to monitor trends and patterns of cancer incidence over time
- to identify high risk populations
- to provide a database for and serve as a resource in conducting epidemiologic studies
- to provide data to assist public health officials, hospital administrators, and physicians to effectively plan services, prioritize health resource allocations and develop and measure prevention and intervention strategies.

History and Funding of the Registry

DCCR was established in 1985 following the introduction of home-rule for the self-governance of the District of Columbia. Prior to that date, the DCCR since 1963 was known as the Metropolitan Washington Cancer Registry with the mandate to gather the occurrence of cancer in the Washington metropolitan area covering Baltimore, Washington and Northern Virginia. The operations of the registry are mandated by Federal Statute code §6.4123 as amended in the Preventive Health Services Amendment Act (D.C. Law 6-83 of 1985, and D.C. Law 8-157 of 1990). Additional funding has been awarded to DCCR from the Centers for Disease Control and Prevention (CDC) with a five-year federal grant aimed at enhancing the timely, complete, and accurate collection of cancer data, and for the computerization of data collection from reporting sources.

Collection of Data

Each District hospital, out-patient surgery center, and pathology laboratory is responsible for the complete ascertainment of all data on cancer diagnoses and treatments provided in its facility within six months of diagnosis. Sources for identifying eligible cases include:

- hospitals
- out-patient surgery centers
- private pathology laboratories
- free-standing radiation centers
- physicians whose cancer cases are not otherwise reported by the facilities listed above
- death certificate sources from State Center for Health Statistics (Vitals Statistics Office)
- out-of-state cancer registries reporting a District resident receiving care in non-District health-care facilities.

When a cancer case is reported from more than one source the information is consolidated into one record. Reported cases contain the following data:

- patient demographics (including geographic place of residence at time of cancer diagnosis)

- description of cancer (including date of diagnosis, primary site, metastatic sites, histology, extent of disease, etc)
- first course of treatment.

Primary site, behavior, grade, and histology are coded according to the *"International Classification of Diseases for Oncology, 2nd Edition"*. Stage of disease variables are coded using *"SEER's Summary Staging Guide"* and *"AJCC Manual for Staging of Cancer, 4th Edition"*. All other variables are coded following the rules of the North American Association of Central Cancer Registries, the SEER program, and the American College of Surgeons.

Reportable Cases

All in-situ or malignant neoplasms are reportable to DCCR, except in-situ cancer of the cervix. The database includes all cases of carcinoma, sarcoma, melanoma, lymphoma, and leukemia, diagnosed by histology/cytology, radiology, laboratory testing, clinical observation, and autopsy.

Basal and squamous cell carcinomas of the skin are excluded except when occurring on a mucous membrane.

In conformance with guidelines provided by the North American Association of Central Cancer Registries, cervix in-situ cases, if reported to the DCCR, have been excluded from this publication.

Confidentiality of Data

The District of Columbia cancer reporting law ensures the protection of confidential data and restricts the release of identifying data. Persons with access to confidential data are required to sign a pledge of confidentiality and are subject to penalty if they, through negligence or willful misconduct, disclose confidential data.

Quality Assurance

To assure validity and reliability of the data presented in this report, DCCR has many mechanisms in place to check data for quality and completeness. We use EDITS software which have standard edits using algorithms that check the content of data fields against an encoded set of acceptable possible contents and flag the acceptability of coded data. Edits include field edits, inter-field edits, and inter-record edits. Edits check for unlikely sex/site, site/histology, or site/age combinations. In addition to computerized edits, each case is manually reviewed for errors.

Records are also routinely checked for duplicate entries. Duplicate case checking is performed both manually and electronically using various methodologies.

The District's 1998 incidence data used for this report have been audited by the North American Association of Central Cancer Registries. NAACCR audits are performed on pre-determined sample of reporting facilities and cases in a state's central cancer registry database for a given year. The procedure involves re-casefinding and re-abstracting studies on selected facilities and cases to determine the extent of completeness and accuracy of cancer data. In the NAACCR Audit of the District's 1998 incidence data, all non-federal reporting hospitals were included in the sample of hospitals, and 10 percent of the reported cases were sampled for re-abstracting studies.

The 1998 incidence data has been submitted to NAACCR for review. This standard setting organization has given the 1998 DC

incidence data a gold certificate performance. This indicates that the data has met the organization's standards for completeness of case ascertainment and of individual data items and accuracy of recorded information.

NAACCR 1993-1997 Incidence and SEER 1993-1997 Incidence and Mortality Rankings:

The latest report on cancer incidence for North America (Cancer in North America 1993-1997, Volume One: Incidence) published by NAACCR presents incidence rates for 50 U.S. states and 12 Canadian provinces. In this report, the incidence rates for the 50 states including D.C. were compared and ranked by incidence rates.

The SEER mortality rankings of states were derived from the National Center for Health Statistics (NCHS) data for the 1993-1997 calendar years. In the comparison of the SEER average mortality for the period 1993-1997, there were several specific cancer sites for which the District was either among the five states with highest or five states with lowest mortality rates compared with the 1998 mortality rates for D.C.

Executive Summary

Data Presentation

This report is comprised of five sections. Section I focuses on the 23 most common sites and all sites combined. Presented are age-adjusted incidence rates, number of cases, number of deaths, counts by ward, stage of disease at time of diagnosis, risk factors, special notes, age-adjusted incidence rate comparisons by ward, age specific rates by ward of residence, 1993-1997 SEER incidence and mortality rates. Also included in Section I are GeoCoded maps of the incidence and mortality rates for the D.C. area. In Section II of the report, the incidence data by site and gender for invasive and in-situ cases are presented. Section III presents the number and age-adjusted incidence rates (mean, median and confidence intervals of the estimated rates) for each of the 8 Wards in the City. Using a similar format as in Sections II and III, the age-adjusted mortality data by site and gender are presented in Section IV, while the number and age-adjusted mortality rates (mean, median and confidence intervals of the estimated rates) for each ward are given in Section V.

Population Description

The population of the District of Columbia in 1998 was estimated to be 551,388, made up of 262,386 males (47.6%), and 289,002 females (52.4%). Although estimates for the Hispanic composition of the District's population in 1996 are unavailable, figures from the 1990 population census indicates that there were a total of 32,710 (5.4%) persons of Hispanic origin counted in that year (606,900). This composition was estimated to have increased to 43,332 (8.1%) Hispanics out of a total population of 535,027 in 1997. Population estimates were obtained from the U.S. Bureau of the Census.

The District of Columbia is comprised of 91 census tracts grouped into eight wards. The total population of each ward is given in the Appendix.

Descriptive Summary by Race and Gender

The data presented in this report cover those cases diagnosed among District of Columbia residents between January 1, 1998 and December 31, 1998. In this time frame, there were 4,201 cases of cancer diagnosed in the District, of which 2,273 (70.1%) were African Americans, 762 (23.5%) were Caucasians, 72 (2.2%) were of other races including Asian Indians/Pakistanis, American Indians/Aleutians/Eskimos, Vietnamese, Chinese and Micronesians, while the remaining 134 (4.1%) were of unknown races. Due to the small number of "Other Races" in the total sample of cases in the database, it was found expedient to combine this group with the Caucasian cases.

Classified by ethnic origin of patients, the cancer cases diagnosed in 1998 consisted of 34 (1.0%) of Hispanics, whereas 2,980 (91.9%) were non-Hispanics, and 226 (7.0%) were of unknown ethnicity. The characteristics of the 1998 cases further indicates that 1,737 (53.6%) of the total 3,259 cases reported were males, while 1,497 (46.2%) were females, and 7 (0.2%) were of

unknown gender. Due to the unavailability of age information for the Hispanic population, and the small number of cases recorded in 1998, rates were not computed separately for this group.

Technical Notes

Age-specific Incidence Rates

Age-specific rates are calculated by dividing the number of cases for a given age-group by the total population of that age-group and are expressed as an average annual rate per 100,000 population by age group. Age-specific rates exclude the same types of cases that are excluded from age-adjusted incidence rates.

Let the age-distribution of deaths occurring in a given geographical area and year be denoted by $\boldsymbol{d_i}$, and let the corresponding distribution of mid-year population at risk be denoted by $\boldsymbol{p_i}$. By definition, therefore, the age-specific death rate for any age interval, m_i, can be expressed as:

$$m_i = \frac{d_i * 100{,}000}{p_i}$$

Eqn. (A)

where *i* denotes each of the *N* categories of age over which the deaths and population are distributed, and the computed rate is expressed per 100,000 population.

Direct Age-Adjusted Death Rate:

In order to permit the comparison of cancer mortality rates across populations whose age distributions might be different, it is recommended that age-adjusted rates using a standard population distribution be obtained. In this report the U.S. Standard Million Population, given in Appendix, Table 21, was employed to perform direct adjustments of the cancer mortality rates presented. The observed deaths and population were grouped into the same eighteen (18) age intervals, 0-4, 5-9, 10-14, ..., 80-84, 85+ as the Standard Million Population. Thus, i = 1, 2, ..., 18, and the age-adjusted incidence or mortality rate (R) can be expressed as:

$$R = \sum_{i=1}^{18} \frac{m_i p_i^s * 100{,}000}{P_T^s}$$

Eqn. (B)

where P^s_T is the sum of the standard population , p^s_i, over all 18 age intervals i.e.

$$P_T^s = \sum_{i=1}^{18} p_i^s$$

Eqn. (C)

The age-adjusted incidence and mortality rates published within this report were adjusted using the direct method and standardized to the age distribution of the U.S. 1970 population (see Appendix for the 1970 U.S. Standard Population). The rate represents the average number of new cases diagnosed annually per 100,000 persons. Age adjustment allows rates from one geographic area to be compared with rates from another geographic area that may have differences in age distributions. Any observed differences in age-adjusted incidence rates between populations are not due to differing age structures.

The computation of rates requires reliable estimates of the population at risk by five-year age groups, gender and race during the time period being studied. Population figures used in this report were obtained from U.S. Bureau of the Census estimates of wards by age, gender and race, as released by the D.C. Office of Planning.

Confidence Interval of Age-Adjusted Death Rate:

The confidence interval of age-adjusted rates gives the estimated range of values within which the true population value lies with given probability. The confidence interval of the age-adjusted rate given in Equation (B) above may be obtained by employing the approximate expression given for the computation of the standard error of a crude mortality or incidence rate (Keyfitz, 1966):

$$C.I._{(R)} = R \pm c_{1-\alpha} S.E._{(R)}$$

Eqn. (D)

where: $$S.E._{(R)} = R / [\text{events}]^{1/2}$$

Eqn. (E)

and events denotes the number of deaths or incident cases upon which the calculation of the rate was based.

Risks and Associated Factors

The "risk and associated factors" subsections in Section I were developed from extracts from the following sources: 1995 Cancer Mortality Report of the District of Columbia, Cancer in Idaho 1996 Annual Report, Cancer Rates and Risks (NIH/NCI 4th Edition, 1996), Cancer Statistics for African Americans 1996, and the web-sites listed in the reference section.

Mean/Median/Mode

Measures of central tendency are helpful to describe a group of individual values in a simple and concise manner:

Mean: also known as the arithmetic average, is the sum of all observations divided by the number of observations.
Median: is the middle value when the observations are ranked in order from the smallest to the largest.
Mode: is the value which occurs most frequently in a group of observed values.

Cancer Case Definition

A "cancer case" is defined as a primary cancer site (where the cancer started), not a metastatic cancer site (where the cancer spread to). Since an individual can have more than one primary cancer site during their lifetime, the number of incident cancer cases are greater than the number of persons who are diagnosed with cancer.

Limitations to Data Interpretation and Comparison

Rates based on population estimates: District-wide and ward population figures are estimates, and errors in the estimates will also impact the rates.

Rates exclude missing/unknown demographic characteristics:
The numerators used in the calculation of age-specific and age-adjusted rates in this report exclude those cases for which the patient's age, gender, or race are unknown or missing. As a result, occasionally tallies for "All Sites" may not equal the sum of the component categories. For example, due to missing gender, the sum of the number of cases for males and females may not sum up to the total number of cases reported for "All Sexes".

Rate comparisons: Age-adjusted incidence and age-specific rates based on a small

number of cases (fewer than 10) may be unstable. In comparing rates among geographic areas (wards or other states), factors such as the absolute numbers of cases and differences in demographic characteristics should be considered. Interpretations without consideration of these factors may be misleading or inaccurate.

Standard Site Analyses Categories

To facilitate interpretation of data and comparisons across registries, DCCR uses standardized groupings of standard site analysis categories. These groupings are consistent with the National Cancer Institute's SEER Program and are adopted by NAACCR. Most neoplasms are grouped by the organ where they occur. Neoplasms of the lymphatic, hematopoietic, as well as reticuloendothelial system are grouped by their histologies (leukemias, lymphomas, etc), and not by the anatomical site where they occurred (see Appendix for groupings of codes).

Stage at Time of Diagnosis

Staging measures the extent of disease at the time of initial diagnosis. Summary staging attempts to group cases with similar prognoses into categories of:

- in-situ (non-invasive)
- localized (cancer confined to the primary site)
- regional (direct extension of tumor to adjacent organs, and/or lymph nodes
- distant (metastasis to tissues or lymph nodes remote from the primary site
- unknown.